# Matching an Evaluation Method to Your Equity Question

*Note: The Equity Evaluation Memo Series is intended to guide OES' commitment to equity in our evaluation process and efforts toward understanding and reducing barriers to equitable access to federal programs. This series is intended to be an internal guidance document for OES team members.*

In the OES Equity Evaluation Series guidance memo on defining equity, we explore the ways in which equity is a nuanced topic that eludes a singular definition. In this memo, we first present a flow chart that can be used to match research questions and data characteristics to potential estimators before presenting several specific examples of research questions matched to estimands and evaluation methods. These evaluation methods generally fall into two categories: descriptive evaluations and impact evaluations. Descriptive evaluations seek to describe observed states of the world; they can tell us which groups were most likely to enroll in a benefits program or what the demographic makeup of a recipient population is, but do not seek to describe program or intervention effects. Impact evaluations do seek to estimate effects of programs or interventions, and may require different types or quantities of data to conduct.

## Defining a Research Question

A targeted, program-specific approach will tailor the formulation of a research question to the specific step in the program where equity is being measured. For example, some programs may have multi-step processes, such as hiring processes or award determinations, that involve applications, screening(s), and selection determinations. In this context, an analyst can choose to evaluate the overall pipeline and where drop off of minority groups occurs, how selected candidates differ from the eligible population, and whether impacts for winning applicants differ by race, among other questions.

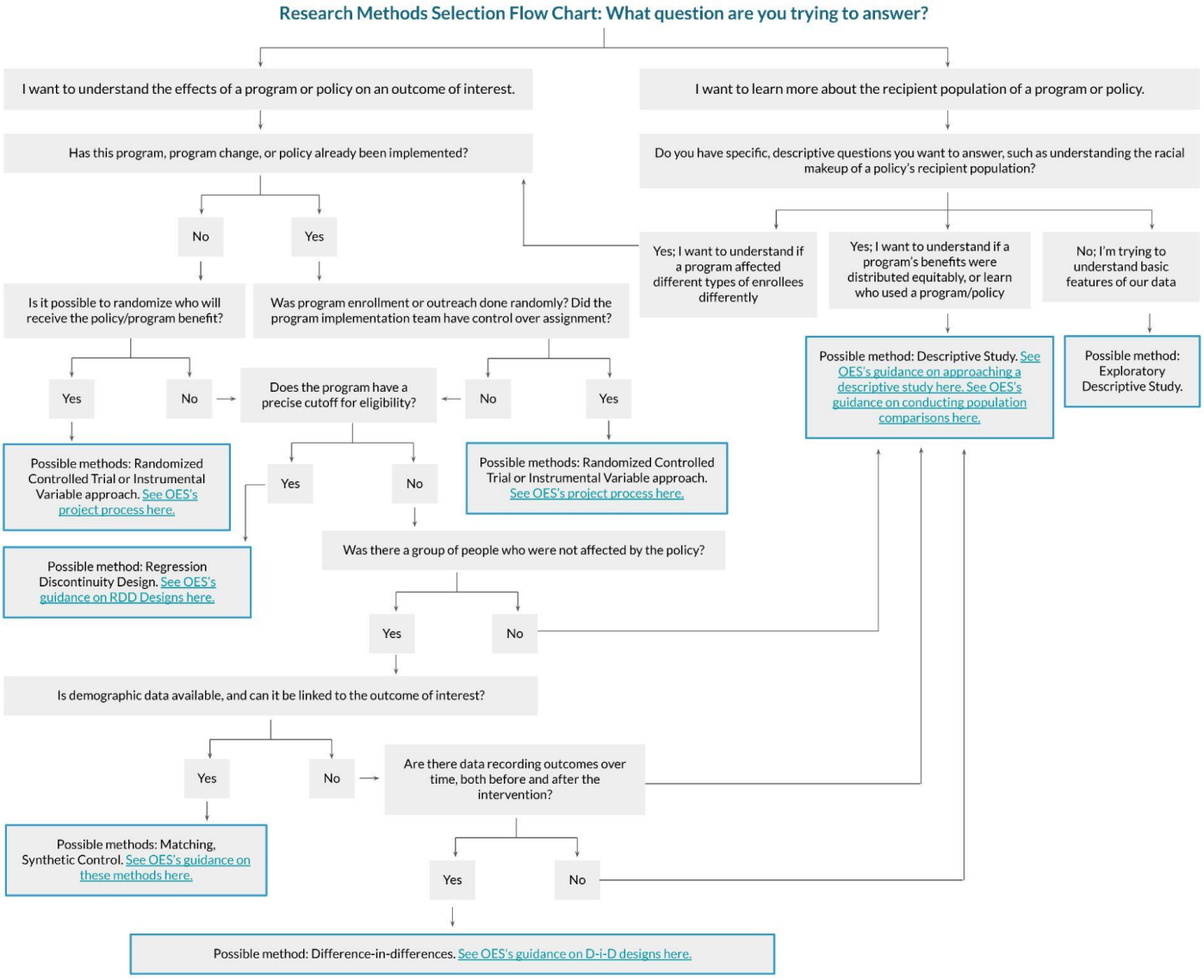
## Choosing an Evaluation Method

Given a research question of interest, analysts should choose a methodology based on a particular understanding of the dimension along which equity is being examined, as well as an understanding of a given program (i.e., the program goals, guidance, and target groups as well as the realities of the implementation) and the available data.

Two broad classes of evaluations OES conducts are descriptive studies and impact evaluations. Descriptive studies describe a program, policy, organization, or population without inferring causality or measuring effectiveness. Descriptive studies can help us understand relationships between program activities and participant outcomes, measure relationships between policies and particular outcomes, describe program participants or components, and identify trends or patterns in data. In contrast, impact evaluations (causal studies) look at outcomes for a given

(change in) program or policy, ideally using a comparison or control group in order to estimate causal impacts. For equity-related work, we note it is particularly important to make the distinction between descriptive and causal studies. The difference is in the research question: descriptive studies do not try to compare counterfactual states of the world to infer causal relationships; causal studies do.

Use the research methods selection flowchart to determine whether your research question is descriptive or causal, as well as to identify a potential methodological approach for evaluation.



## Examples:

The following examples illustrate how certain methods are best suited for answering particular equity questions.

- **Research question:** What is the impact of a change to a program or policy on who participates?

  *Method:* OES used a quasi-experimental design to look at the impact of a change to how applications for a grant program were processed, on the percent of applicants who came from minority backgrounds. We found that, when the application process was simplified, the percent of applicants from minority backgrounds increased.

- **Research question:** What are the impacts of an intervention among different groups? Is there a significant difference between these impacts?

  *Method:* Early research on Head Start, an early childhood intervention, showed few lasting effects on most outcomes of interest. However, when subsequent researchers looked at the impacts of Head Start on different subgroups, they found that this small average effect hid a wide range of effects on these groups (for example, Kline and Walters, 2016). For instance, they found that children with lower cognitive skills benefited most from Head Start, but those with special needs may have experienced negative effects. OES used a randomized controlled trial with pre-specified subgroup analysis to show that text reminders were more effective among women under age 25 than women as a whole in increasing health clinic visits in Mozambique.

- **Research question:** What proportion of people have a given attribute or set of attributes, at a given stage in a program? How does this compare to what we expect or hope to see?

  *Method:* OES conducted a descriptive study examining whether the demographics of families who received Emergency Rental Assistance look similar to demographics of families the ERA was intended to help. Black renters were strongly overrepresented among recipients of ERA – their share of the recipient population was 21-22 percentage points higher than their share of the eligible population.

- **Research question:** Do key outcomes differ for different groups?

  *Method:* Suppose an NGO that provides job training to formerly incarcerated people is interested in whether people who participate in its program have different experiences when looking for jobs, depending on their race. A descriptive study seeking to answer this question could track participant's employment rates and income for a few years after they leave the program, and report trends for Black, white, and Hispanic participants.

The following table summarizes these examples, linking research questions to a description of the relevant estimand and potential evaluation methods. Please note that some of the notation and guidance linked in this table is fairly technical, and may not be useful for all readers.

| **Research Question** | **Description** | **Possible estimation approaches and examples** |
|---|---|---|
| *What is the average treatment effect for a particular group?* | E[Y(T = 1) - Y(T = 0)]<br><br>The expected difference in potential outcomes within a population of interest. | RCTs as well as quasi-experimental designs |
| *Does an intervention have unequal impacts on certain subgroups?* | E[Y(T = 1) - Y(T = 0) \| A = a] - E[Y(T = 1) - Y(T = 0) \| A = a']<br><br>This estimand describes whether the average causal effect is larger or smaller among different subgroups. This is sometimes called causal moderation analysis. For example, one might want to know whether an intervention to increase application to ERA is stronger for low-income people than for non-low-income people. | In experimental settings, estimators can be very simple (difference in differences-in-means, multiple regression with interaction). In quasi-experimental settings, may require a more complicated estimation strategy. |
| *What is the average treatment effect of class membership on outcomes?* | E[Y(A = a) - Y(A = a') \| B = b]<br><br>The expected difference in potential outcomes brought about by changing someone's membership in a class while holding other features constant. | Analyses that try to examine class-based discrimination by looking at whether "differences persist" after controlling for all other relevant factors. Oaxaca blinder and multiple regression, are examples. |
| *What proportion of people have a given attribute or joint set of attributes, at a given stage in the program?* | E[A = a \| B = b]<br>Pr[A = a, B = b \| Y = y]<br><br>This descriptive estimand assesses what proportion of people have a given attribute or joint set of attributes, given some stage in the program. Can be useful on its own for process evaluation. | In cases with good data, can be achievable through simple conditional means. In cases with less rich data, may require estimating models and extrapolating, ecological inference, and so forth. |
| *Do the recipients of a program resemble those who were its eligible or intended recipients?* | Pr(A = a \| Y = y, B = b) - Pr(A = a \| Y = y', B = b)<br><br>This estimand assesses the degree to which a group that normatively *should* have some degree of access to a program or receive some share of its benefits actually *does*.<br><br>Suppose, for example, that A = a means an applicant is a woman, B = b means they need ERA, Y = y indicates they received ERA, and Y = y' indicates they did not. Then this estimand calculates the | Conditional means for point estimation, multinomial tests or F-tests for hypothesis testing. |

| | | |
|---|---|---|
| | degree to which women in-need of ERA were represented among the actual recipients. Note that the group for which Y = y can be much larger than the one for whom Y = y', and vice versa. The key is that we are interested in the proportion occupied by women in both groups.<br><br>Similarly, one could imagine that B = b indicates having applied for ERA, Y = y indicates one's case was selected for review, and Y = y' indicates that it wasn't. Then this estimand describes the rate at which women drop out of the application process relative to others. | |